Forecasting northern polar stratospheric variability using a hierarchy of statistical models

by

Ivan Minokhin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The northern polar stratosphere plays an important role in modulating the wintertime near-surface temperature conditions in midlatitudes. Forecasting northern polar stratospheric variability will have the potential to extend the winter weather forecasts in midlatitudes. As such, this research seeks to explore a novel approach of forecasting short-term northern polar stratospheric variability using a hierarchy of linear and non-linear statistical models. In addition to El Niño Southern Oscillation, the Quasi-biennial Oscillation, and the 11-year solar cycle indices, this research uses the upward flux of wave activity from the troposphere into the stratosphere as a predictor for modeling and forecasting northern polar stratospheric temperature and geopotential height anomalies. The upward flux of wave activity entering the stratosphere is the primary source of intraseasonal variability in the wintertime stratospheric polar vortex. Multiple linear regression and machine learning models were trained over the 1980-2005 time period, and the 10-day and 20-day northern polar stratospheric temperature and geopotential height forecasts were generated over the 2005-2011 time period. The importance of each predictor for modeling northern polar stratospheric variability was assessed using a permutation-based method. The study has found that the use of the meridional wave heat flux predictors improves the accuracy of short-term northern polar stratospheric geopotential height forecasts as demonstrated by the correlation coefficient of 0.48 over the 2005-2011 time period. In contrast to previous studies, multiple linear regression shows better predictive performance than the machine learning models over the 2005-2011 time period. A better predictive performance of multiple linear regression in comparison to machine learning models is due to a much higher contribution of the upward flux of wave activity than other predictors to forecast skill.

Acknowledgements

I would like to thank all the people who made this thesis possible. Dr.Chris Fletcher for his thoughtful suggestions and continued advice. Dr.Alexander Brenning for his willingness to share his expert knowledge of statistical modeling. Lastly, I am most grateful to my family for their support and understanding.

Preface

This thesis contains the completed manuscript (Chapters 1,2,3 and 4) to be submitted to *Climate Dynamics* journal. The manuscript explores the short-term predictability of the northern polar stratospheric variability. I completed all analysis and generated northern polar stratospheric forecasts with statistical models using statistical software *R*, then wrote the manuscript. This is the first research to use the upward propagating flux with wavenumber 1 and wavenumber 2 entering the stratosphere as predictors for forecasting northern polar stratospheric variability using statistical models. In addition to multiple linear regression and machine learning models such as artificial neural networks and support vector regression, random forest was also employed for the first time to generate forecasts of northern polar stratospheric temperature and geopotential height anomalies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Forecasting northern polar stratospheric variability has the potential to improve the accuracy of the Northern Hemisphere tropospheric winter temperature forecasts and as such provide socio-economic benefits (Sigmond et al., 2013). The northern polar stratospheric variability represented by the polar cap average of temperature and geopotential height anomalies at 10hPa was modelled and forecasted with linear and non-linear statistical models using climate indices representing the QBO, ENSO, solar variability, and 40-day averaged meridional wave heat fluxes as predictors. Multiple linear regression and machine learning models were trained over the time period between July 1, 1980 and June 30, 2005 and the 10-day and 20-day forecasts were generated for the time period between July 1, 2005 and June 30, 2011. The model predictive performance was assessed based on $RMSE$ and the correlation coefficient ρ. A permutation-based method was conducted to assess the importance of each predictor. The use of the meridional wave heat fluxes improves the quality of short-term forecasts. Multiple linear regression outperforms machine learning models in forecasting variability over the 2005-2011 time period.

1.2 Background

The variability of the northern polar stratosphere influences wintertime midlatitude near-surface temperature conditions (Thompson et al., 2002; Gerber et al., 2009). In wintertime,

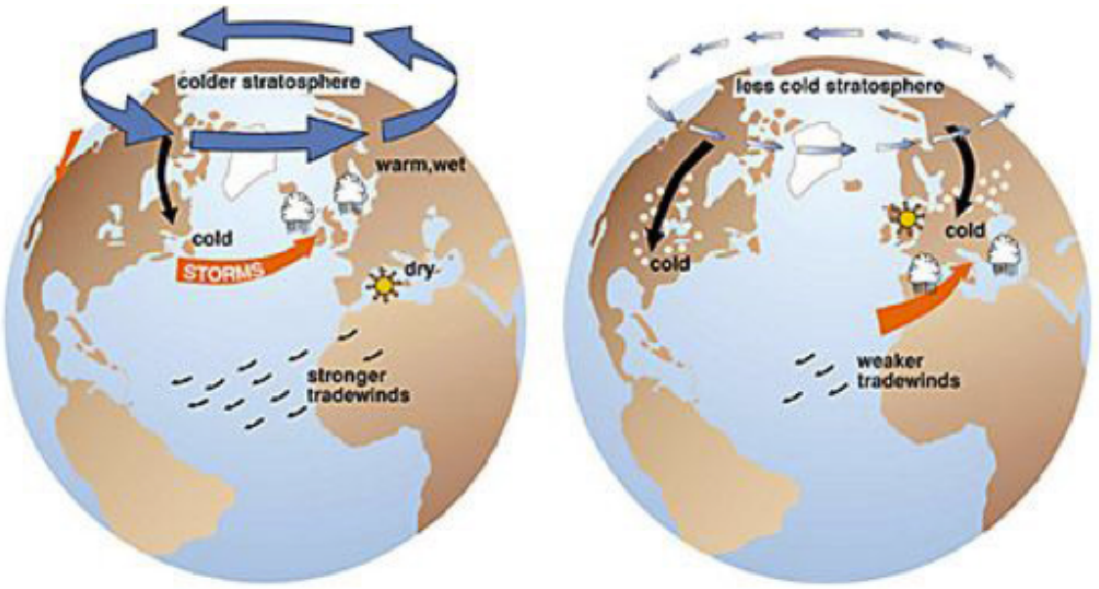

Figure 1.1: The impact of the northern polar stratospheric variability on the weather patterns in the midlatitudes. Left: a stronger polar vortex. Right: a weaker polar vortex. Source: https://www2.ucar.edu/for-staff/update/cycles-dipoles-and-oscillations

the northern polar stratosphere is dominated by the stratospheric polar vortex circumscribed by the stratospheric polar jet (Figure 1.1). A stratospheric polar vortex slowdown and a stratospheric polar jet weakening are associated with a stratospheric polar temperature and geopotential height increase over the period of few days (Baldwin and Dunkerton, 2001; Polvani and Waugh, 2004). Such a phenomenon is known as sudden stratospheric warming (SSW), following which cold Arctic air descends and moves south resulting in cold temperature outbreaks in the Northern Hemisphere midlatitude troposphere. It takes up to 60 days for the strong stratospheric anomalies to propagate to the Earth's surface (Baldwin and Dunkerton, 2001). Therefore, northern polar stratospheric forecasts can lead to improvements in the seasonal climate forecasts (Tripathi et al., 2014), which in turn can have many socio-economic benefits (Sigmond et al., 2013). One such impact of a colder than average winter near-surface temperature in the midlatitudes is a demand for energy (Ouzeau et al., 2011). As such, reliable northern polar stratospheric forecasts can be of particular value to energy companies, electricity distributors, reinsurance companies and hedge funds trading weather derivatives (Alaton et al., 2002; Hull et al., 2013).

An approach by Blume and Matthes (2012) (henceforth BM12) indicated that fore-

casting northern polar stratospheric temperature and geopotential height anomalies with statistical models based on the ERA-Interim reanalysis dataset, the global atmospheric reanalysis dataset composed by combining model simulations and observations (Dee et al., 2011), was moderately successful with statistical models, particularly artificial neural networks. The statistical modeling in BM12 involved partitioning the dataset into the training set and the test set. Having trained statistical models over the time period between July 1, 1980 and June 30, 2005 and generated forecasts for the test set between July1, 2005 and June 30, 2011, BM12 concluded that for forecasting geopotential height anomalies artificial neural networks had better predictive performance than multiple linear regression as shown by the correlation coefficient between actual and forecasted values over the test set: 0.42 from artificial neural networks and 0.31 from multiple linear regression. Moreover, BM12 demonstrated that for modelling temperature and geopotential height anomalies, predictors such as the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation representing sea surface temperature (SST) variability, high latitude blockings, and aerosol optical depth representing volcanic activity, were less important than the Quasi-biennial Oscillation (QBO), El Niño Southern Oscillation (ENSO) and the 11-year solar cycle.

Though achieving moderate success with forecasting 2011/2012 wintertime conditions and highlighting the importance of the QBO, ENSO and the 11-year solar cycle as predictors, BM12 recognized that modeling northern stratospheric variability, especially temperature, needed to be improved and identified a need for additional predictors. Furthering the work completed by BM12, this research will add meridional wave heat fluxes as predictors.

Northern polar stratospheric variability is influenced by the planetary-scale waves that propagate from the troposphere (Mukougawa and Hirooka, 2004; Tripathi et al., 2014). As these planetary waves enter the stratosphere, they break and deposit their energy resulting in a warmer polar stratosphere, a weaker stratospheric polar jet, and a higher likelihood of a SSW occurrence. SSW has been shown to be influenced by the Eliassen-Palm flux, which in turn is proportional to the meridional wave heat flux (Andrews et al., 1987). The meridional wave heat flux with wavenumbers 1 and 2 at 100hPa, have been shown to influence the state of the polar stratosphere in the Northern Hemisphere (Newman et al., 2001; Polvani and Waugh, 2004; Nakagawa and Yamazaki, 2006; Charlton and Polvani, 2007; Fletcher and Minokhin, 2015). Seeing as heat fluxes strongly influence the intraseasonal stratospheric variability, the use of heat fluxes as predictors is likely to lead to an improved accuracy of short-term northern polar stratospheric temperature and geopotential height forecasts obtained from statistical models. As such, heat fluxes are used as predictors in this research.

Furthermore, the meridional wave heat flux can be decomposed further into the linear

(LIN) and non-linear (NONLIN) terms representing wave interference between the climatological stationary waves and the anomalous waves (Smith and Kushner, 2012). The LIN term representing a phasing effect or linear interference was shown to be the dominant term (Smith and Kushner, 2012; Fletcher and Minokhin, 2015) and thus the LIN terms of wavenumber 1 and 2 (LINw1 and LINw2) are also included as predictors in this research. The importance of such predictors for modeling stratospheric polar variability will be assessed in comparison to the QBO, ENSO, and 11-year solar cycle predictors used by BM12. Including both the TOTAL and LIN heat fluxes as predictors will allow for comparison of the importance of the TOTAL and LIN meridional wave heat fluxes as predictors.

In addition to using the meridional wave heat fluxes to improve the quality of short-term northern polar stratospheric temperature and geopotential height forecasts, the proposed approach also aims to address two limitations of the approach by BM12. First, BM12 used extensively the principal component analysis to determine temperature and geopotential height response variables, as well as predictors influencing northern polar stratospheric variability. Computing the leading principal components from daily data over a 25-year time period with many grid points in the ERA-Interim dataset is computationally expensive. Moreover, the principal components are sensitive to the time period over which they are calculated. This research avoids the use of the principal component analysis in computing both the response and predictors and, as a result, the proposed approach is much simpler and the results are expected to be easier to reproduce and to interpret.

Second, BM12 generated 2011/12 northern polar stratospheric temperature and geopotential height forecasts based on a combination of lagged predictors, SST predictions from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center, as well as subjective assumptions about predictors' future values. Such approach may invalidate the accuracy of forecasts since it is unclear if those assumptions will hold true in the future. The proposed approach aims to generate northern polar stratospheric forecasts based entirely on the past values of predictors. The optimal lags between the response variables and predictors will be determined and a minimum lag of 10 and 20 days will be introduced to ensure that the forecasts can be generated 10 and 20 days ahead. The duration of the forecasts in this research is similar to the duration (5-14 days) of the forecasts obtained from general circulation models (GCMs) (Tripathi et al., 2014).

1.3 Motivation for research

The importance of the northern polar stratosphere in modulating near surface weather conditions is highlighted by the fact that after Environment Canada was able to increase

the stratospheric resolution in their forecast model by raising the model lid from 10hPa to 0.1hPa, a resulting large improvement in accuracy of stratospheric forecasts led to a higher accuracy of tropospheric forecasts (Charron et al., 2012). Although forecasting northern polar stratospheric variability can be performed with general circulation models (GCMs), such models suffer from systematic biases due to parameterization of physical processes (Kerkhoff et al., 2014; Wang et al., 2014). Additionally, GCM simulations exhibit weaker Northern Hemisphere stratospheric polar variability than observations and consequently extreme events such as sudden stratospheric warming (SSW) may not be captured by model simulations (Charlton et al., 2007; Sheshadri et al., 2015).

The use of statistical models for forecasting variability is a good alternative to computationally intensive GCMs. Statistical models are simpler, computationally more efficient, and can provide an insight about forcing interactions (Blume and Matthes, 2012; Deo and Şahin, 2015). In particular, machine learning models are well suited to model the complex non-linear forcing interactions (Hsieh, 2009) and they have been successfully applied to the field of environmental science such as separating linear and nonlinear El Niño teleconnections using neural networks (Hsieh et al., 2006) and forecasting winter extreme precipitation using support vector regression (Zeng et al., 2011). In this research the accuracy of forecasts from machine learning models will be compared against multiple linear regression.

To effectively forecast temperature and geopotential height anomalies using statistical models, the factors influencing northern polar stratospheric variability need to be identified. Among such factors that have been identified are El Niño Southern Oscillation (ENSO) (Garfinkel and Hartmann, 2008; Ineson and Scaife, 2009; Butler and Polvani, 2011; Garfinkel et al., 2012), the Quasi-biennial Oscillation (QBO) (Lu et al., 2008; Naoe and Shibata, 2010; Richter et al., 2011; Watson and Gray, 2014) and the 11-year solar cycle (Gray et al., 2010; Smith et al., 2012). This research will also incorporate other predictors of intraseasonal stratospheric variability. Planetary-scale waves of wavenumber 1 and wavenumber 2 propagate from the troposphere and contribute to a polar stratospheric warming and a geopotential height increase (Mukougawa and Hirooka, 2004; Stan and Straus, 2009; Tripathi et al., 2014). The meridional wave heat flux with wavenumber 1 and 2 entering the stratosphere were shown to influence northern polar stratospheric variability (Newman et al., 2001; Polvani and Waugh, 2004; Nakagawa and Yamazaki, 2006; Charlton and Polvani, 2007; Fletcher and Minokhin, 2015) and therefore including such heat fluxes as predictors may lead to higher accuracy of northern polar stratospheric forecasts. Additionally, since the meridional wave heat fluxes can be decomposed into linear and non-linear components (Smith and Kushner, 2012; Fletcher and Minokhin, 2015), the use of linear components as predictors will be assessed as well.

Forecasting northern polar stratospheric variability can be used to assess the weather risk associated with below average wintertime midlatitude temperature conditions. Strong northern polar stratospheric wintertime anomalies increase a likelihood of very cold temperature outbreaks in the midlatitudes. One such impact of a higher number of cold winter days is a higher demand for heating sources and therefore reliable forecasts will be of value to energy companies, electricity providers, reinsurance companies and weather derivative traders. As an example, the electricity cost in the province of Québec is lowest in Canada and, as a result, most households use electricity to heat houses in winter. An increased electricity demand in wintertime and the resulting pressure on the Hydro-Québec power grid prompted Hydro-Québec to offer the residential DT rate[1], which features a five fold increase in electricity price when the temperature is below -12 °C . The DT rate is a classic example of weather derivatives known as heating degree days (HDDs) (Hull et al., 2013). Since Hydro-Québec customers can lock for a year into either a regular tiered rate (D rate) or DT rate at any time throughout the year, the knowledge of a higher likelihood of cold winter days can assist in either electricity rate selection, or a better preparation for reduced electricity consumption during those cold winter days. Though the savings from switching between the D and DT rates may not be substantial, the reliable northern polar stratospheric forecasts can aid in a better planning and decision making, especially if losses associated with the cold outdoor temperature in wintertime are large.

1.4 Research objectives

The goal of this research is to improve short-term northern polar stratospheric temperature and geopotential height forecasts based entirely on lagged predictors using competing statistical models. In addressing this goal, this research seeks to explore the following objectives in an effort to better understand the importance of predictors and the use of statistical models for forecasting northern polar stratospheric variability.

- To explore the importance of meridional wave heat fluxes as predictors in forecasting northern polar stratospheric variability as opposed to ENSO, the QBO and the 11-year solar cycle;
- To determine whether machine learning models exhibit higher predictive performance than multiple linear regression in forecasting northern polar stratospheric variability;

[1] http://www.hydroquebec.com/residential/understanding-your-bill/rates/residential-rates/rate-dt/

- To investigate the sensitivity of the accuracy of forecasts with respect to time lags and to test whether there is a deterioration in the accuracy of the 20-day forecasts as opposed to the 10-day forecasts;
- To investigate if SSW, which is an example of extreme polar stratospheric events, can be forecasted with statistical models using the proposed set of predictors.

1.5 Thesis structure

This thesis consists of four chapters. Chapter 1 describes the role of the northern polar stratospheric variability in modulating wintertime tropospheric temperature conditions in the midlatitudes. Prior research in forecasting northern polar stratospheric variability is discussed and a new approach is presented. Chapter 1 (sections 1.2 and 1.4), and chapters 2,3 and 4 contain the manuscript of a journal article. The manuscript outlines the process of generating short-term temperature and geopotential height forecasts with statistical models using meridional wave heat fluxes as predictors. Chapter 2 describes data, the response and predictors, linear and non-linear statistical models, and methods for data scaling and predictor importance assessment. Chapter 3 summarizes the regression and forecast results, as well as the importance of each predictor in generating regression results. Chapter 4 contains the discussion of the results, describes the limitations in generating forecasts and provides suggestions for future research.

Chapter 2

Data and Methods

This section outlines the data, response variables and predictors, provides an overview of statistical models, as well as methods for data scaling, predictive performance assessment and the importance of each predictor.

2.1 Response variables

The ERA-Interim dataset at 0.75 °x 0.75 ° grid resolution was used to obtain daily northern polar stratospheric temperature and geopotential height response variables, as well as predictors influencing northern polar stratospheric variability. Response variables in this research are the polar cap average of the daily stratospheric temperature (T) and geopotential height (Z) anomalies at 10hPa. The time series of response variables over the time period between 1979 and 2015 are shown in Figure 2.1. The daily time series of T and Z are characterized by much higher fluctuations in wintertime than in the summer. The very large wintertime peaks in both T and Z time series are associated with SSWs when both temperature and geopotential height increase rapidly within a time span of few days. Principal component analysis was not employed in this research to compute the response variables as in the previous study by BM12, who defined the response variables as the leading principal component of the polar cap average of temperature (P1T) and geopotential height (P1Z) anomalies at three levels: 10hPa, 20hPa and 30hPa. Response variables T and Z are highly correlated with P1T and P1Z with the correlation coefficient greater than 95% and therefore the regression and forecast results in this research can be compared to BM12 results.

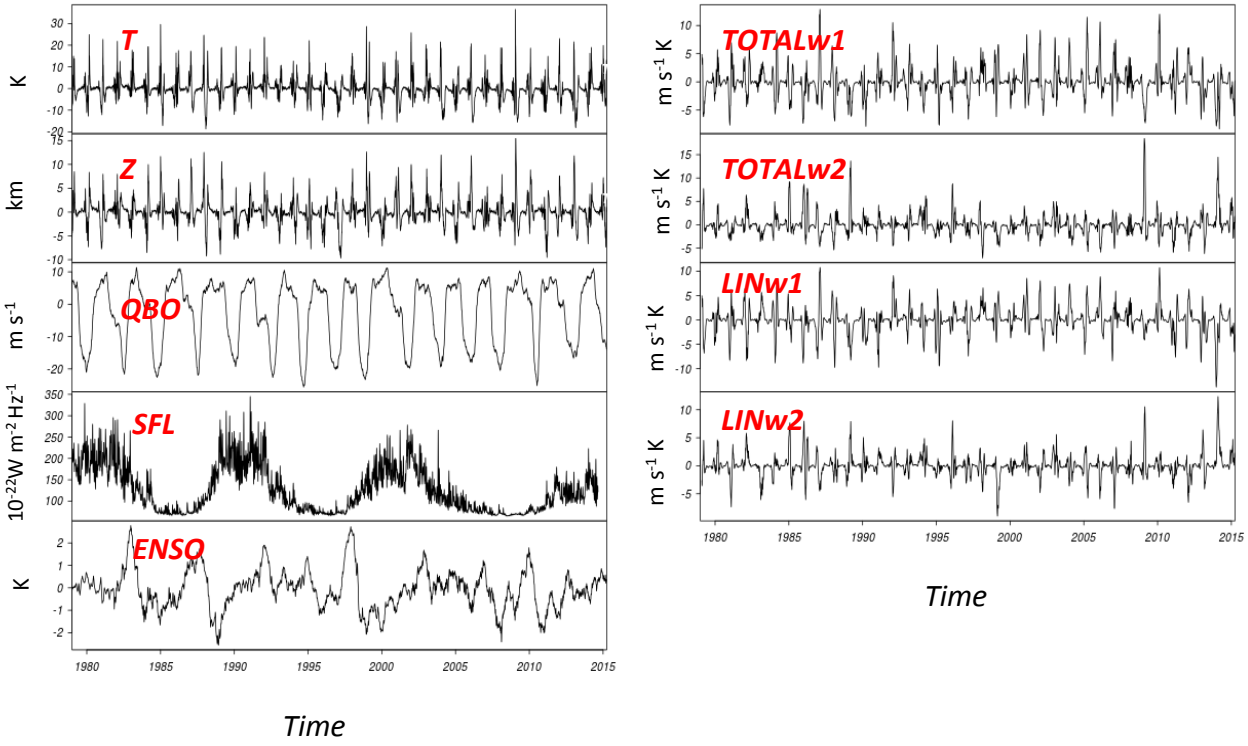

Figure 2.1: Time series of temperature (T) (units in K) and geopotential height (Z) (units in km) at 10hPa, QBO (units in ms^{-1}), SFL (the 11-year solar cycle) (units in $10^{-22}Wm^{-2}Hz^{-1}$), ENSO (units in K) and the total meridional wave heat flux of wavenumber 1 and 2 (TOTALw1 and TOTALw2), as well as the linear component of the total meridional wave heat flux of wavenumber 1 and 2 (LINw1 and LINw2) (units in $ms^{-1}K$).

2.2 Predictors

The predictors are climate indices representing the QBO, ENSO, the 11-year solar cycle, as well as the TOTAL and LIN heat fluxes with wavenumbers 1 and 2. The predictor time series are presented in Figure 2.1. The ENSO index was computed as the area-average of the SST anomalies over the Nino3.4 region (Trenberth, 1997). The QBO index was defined as the area-average of the zonal wind at 50hPa between 10 °S and 10 °N latitudes (Garfinkel and Hartmann, 2008). As in BM12, SFL, which is F10.7cm radio flux[1], repre-

[1]ftp://ftp.ngdc.noaa.gov/STP/SOLAR DATA

sents the 11-year solar cycle. Same as in BM12, the QBO, ENSO and SFL predictors in this research were smoothed with a 5-day running mean to lessen the impact of extreme observations. Both the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation representing long-term SST variability were not used because of a small predictive ability of such predictors as demonstrated by BM12. Instead of the leading principal components of the area-average of geopotential height anomalies at 500hPa representing high latitude blockings, the predictors representing the upward heat flux entering the stratosphere such as meridional wave heat fluxes of wavenumber 1 and 2 were used. Finally, no supplementary predictors such as a linear trend and the sine and cosine functions with a period of one year representing seasonality were used. Instead, both the response variables and predictors in this research were deseasonalized by removing the climatological daily mean values computed over the 1980-2005 time period.

Meridional wave heat fluxes of wavenumber 1 and 2 were determined by wave decomposition of the product of $\{v^*T^*\}$, where v^* denotes eddy meridional wind, and T^* denotes eddy temperature at 100hPa, and $\{.\}$ represent the area-average between 45 °N and 75 °N (Fletcher and Minokhin, 2015). The LIN component of the total meridional heat flux was computed as

$$LIN = \{v_c^* T'^* + v'^* T_c^*\}$$

where v_c^* and T_c^* are the climatological daily eddy meridional wind and temperature means computed over the 1980-2005 time period, and v'^* and T'^* are eddy meridional wind and eddy temperature anomalies. Furthermore, meridional wave heat fluxes were averaged over 40 days prior to the date (Smith and Kushner, 2012).

2.3 Optimal lags

To generate 2011/12 wintertime forecasts BM12 used a combination of lagged predictors, NOAA SST predictions and making assumptions as to what the future values of the predictors would be. Though such approach enables one to produce the seasonal climate forecasts, the accuracy of such forecasts remains unclear. This research aims to generate forecasts solely based on the predictors' past values. Although the time lags can be based on the desired forecast horizon, say 30 days or 60 days, such time lags may not be optimal and may lead to poor regression results. As in BM12, the optimal lags between predictors and response variables were determined based on the largest correlation, irrespective of the sign, between the response variables and lagged predictors over the time period between July 1, 1980 and June 30, 2005 (Figure 2.2 and Figure 2.3). However, in contrast to BM12, in this research the lags do not vary from 0 to 365 days, but instead from l to 365 days,

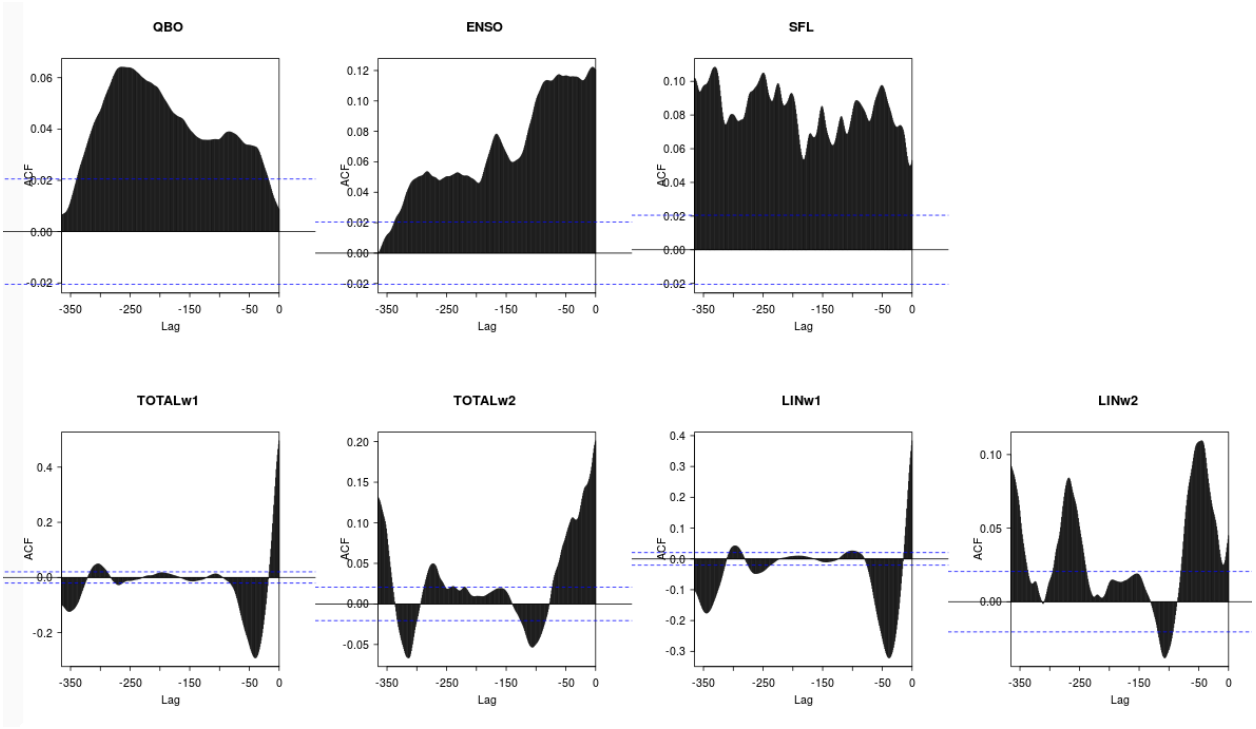

Figure 2.2: The correlation between the temperature response variable and lagged predictors with lags varying from 0 to 365 days. The time period is between July 1, 1980 and June 30, 2005.

where l is a minimum lag. The minimum lag was introduced to ensure that the forecasts can be generated l days ahead based on observations. The minimum lag l in this research is 10 days and 20 days and consequently the 10-days and 20-days forecasts were generated. Such duration of forecasts is comparable with the forecast duration between 5 to 14 days from GCM (Tripathi et al., 2014). The optimal lags between the response variables and predictors are presented in Table 2.1 While the optimal lags for the QBO and SFL remained the same when the minimum lag of either 10 days or 20 days was introduced, the optimal lags for ENSO and meridional wave heat fluxes varied in response to different minimum lags. Such sensitivity of heat fluxes to a minimum lag emphasizes a short memory of heat flux time series. The cross-correlation matrix for the response variables and lagged predictors subject to a minimum lag of 10 days is presented in Table 2.2. The correlation between the temperature and geopotential height response variables is 0.69

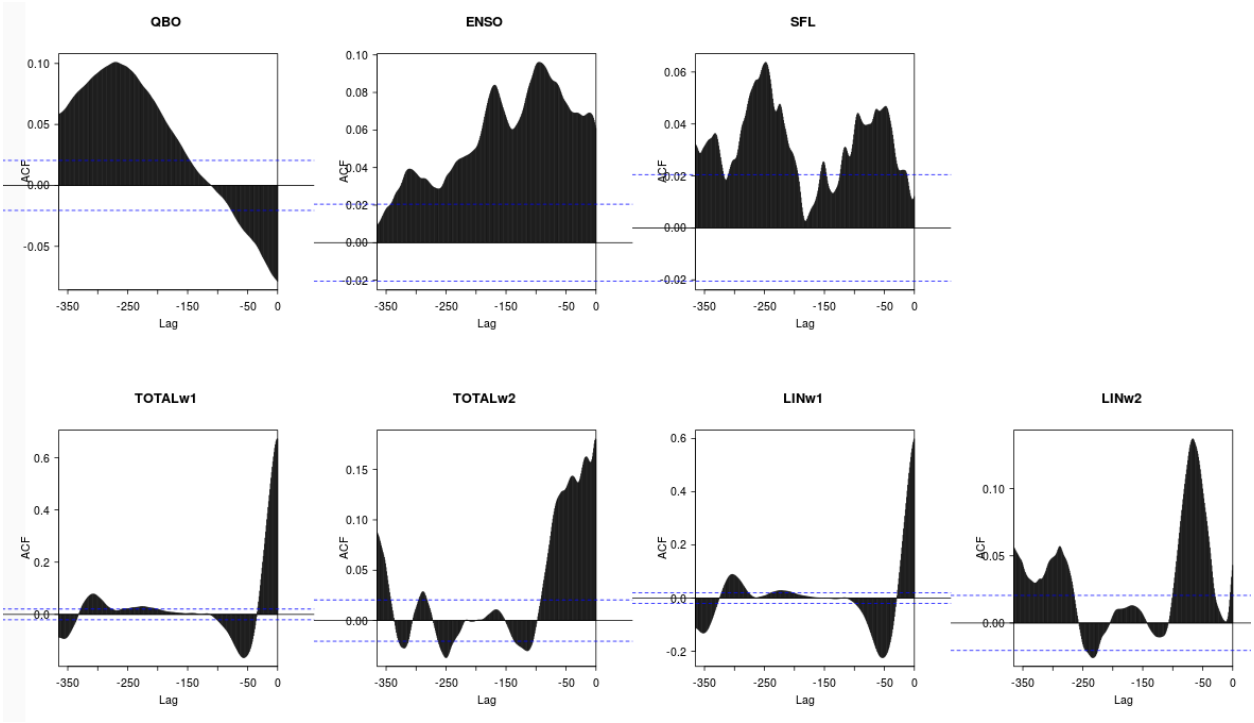

Figure 2.3: Same as in Figure 2.2, but for the geopotential height response variable.

Table 2.1: The optimal lags (in days) between response variables and predictors subject to a minimum lag of 10 and 20 days.

Response variables	Temperature		Geopotential	
Minimum lag	10	20	10	20
QBO	264	264	270	270
ENSO	11	62	95	95
SFL	331	331	248	248
TOTALw1	40	40	11	21
TOTALw2	11	21	16	21
LINw1	39	39	11	53
LINw2	45	45	46	46

Table 2.2: The cross-correlation matrix for (a) temperature (T) and (b) geopotential height (Z) response variables and lagged predictors subject to a minimum lag of 10 days. The time period is between July 1, 1980 and June 30, 2005.

a) Temperature

	T	ENSO	QBO	SFL	TOTALw1	TOTALw2	LINw1	LINw2
T	1.00	0.12	0.07	0.09	-0.29	0.15	-0.32	0.11
ENSO	0.12	1.00	-0.19	0.12	0.13	-0.09	0.09	-0.02
QBO	0.07	-0.19	1.00	0.02	0.04	0.13	0.02	0.18
SFL	0.09	0.12	0.02	1.00	0.03	0.01	0.01	0.05
TOTALw1	-0.29	0.13	0.04	0.03	1.00	-0.19	0.87	-0.38
TOTALw2	0.15	-0.09	0.13	0.01	-0.19	1.00	-0.22	0.13
LINw1	-0.32	0.09	0.02	0.01	0.87	-0.22	1.00	-0.38
LINw2	0.11	-0.02	0.18	0.05	-0.38	0.13	-0.38	1.00

b) Geopotential height

	Z	ENSO	QBO	SFL	TOTALw1	TOTALw2	LINw1	LINw2
Z	1.00	0.10	0.11	0.05	0.52	0.16	0.42	0.09
ENSO	0.10	1.00	-0.17	0.02	0.19	-0.14	0.18	-0.07
QBO	0.11	-0.17	1.00	0.03	0.05	0.13	0.05	0.18
SFL	0.05	0.02	0.03	1.00	0.03	0.00	0.01	0.06
TOTALw1	0.52	0.19	0.05	0.03	1.00	-0.32	0.87	-0.04
TOTALw2	0.16	-0.14	0.13	0.00	-0.32	1.00	-0.32	0.21
LINw1	0.42	0.18	0.05	0.01	0.87	-0.32	1.00	-0.09
LINw2	0.09	-0.07	0.18	0.06	-0.04	0.21	-0.09	1.00

2.4 Data scaling

All predictors were scaled over both the training set and test set such that the minimum of the time series was -1 and the maximum was +1. Such scaling is meant in particular to improve the performance of statistical models, namely artificial neural networks (Hastie et al., 2009), and is consistent with the method used by BM12 to scale the predictors for modeling temperature and geopotential height anomalies. The scaled values $\hat{x}_i$ of the predictor x were computed according to the formula

Table 2.3: The list of acronyms.

Name	Definition
ANN	Artificial neural network
ENSO	El Niño Southern Oscillation
GCM	General circulation model
LINw1	Linear component of the total meridional wave heat flux of wavenumber 1
LINw2	Linear component of the total meridional wave heat flux of wavenumber 2
MLR	Multiple linear regression
MSE	Mean square error
QBO	Quasi-biennial Oscillation
RF	Random Forest
RMSE	Root mean square error
RSS	Residual sum of squares
TOTALw1	Total meridional wave heat flux of wavenumber 1
TOTALw2	Total meridional wave heat flux of wavenumber 2
SFL	11-year solar cycle
SST	Sea surface temperature
SSW	Sudden stratospheric warming
SVR	Support Vector Regression

$$\hat{x_i} = 2\frac{x_i - x_{min}}{x_{max} - x_{min}} - 1 \qquad \forall i = 1, ..., n$$

where x_{max} (x_{min}) is the maximum (minimum) value of an predictor x, and n is the total number of values. Statistical software R was used to complete all calculations (RCoreTeam, 2015) and the list of acronyms for key terms was summarized in Table 2.3

2.5 Statistical models

Multiple linear regression and three machine learning models (random forest, artificial neural networks, and support vector regression) were employed to model and forecast stratospheric polar temperature and geopotential height anomalies.

2.5.1 Multiple linear regression

Multiple linear regression (MLR) (James et al., 2013) assumes that response variable Y can be modelled as a linear combination of predictors X_{j} such that

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_{\mathrm{j}} X_{\mathrm{j}} + \epsilon$$

where β_j are regression coefficients, p is the number of predictors and ϵ is the error term (James et al., 2013). The MLR coefficients are determined by applying the method of least squares which involves minimizing the residual sum of squares (RSS) where

$$RSS = \sum_{j=1}^{p} (y_{\mathrm{i}} - \beta_0 - \beta_{\mathrm{j}} X_{\mathrm{ij}})^2$$

MLR has been used to model stratospheric variability (Randel et al., 2009; Gebhardt et al., 2014). Although the linearity and additivity assumptions of MLR are quite restrictive, MLR is used in this research as the principal model due to its simplicity and a lower computational cost in comparison to machine learning models. Additionally, MLR forecast skill is used as a benchmark for comparing the forecast skill from machine learning models.

2.5.2 Random forest

Random forest (RF) belongs to a class of the tree-based ensemble models and has been shown to be effective at modelling non-linear data (Breiman, 2001). The random forest predictions are obtained by training a large number of regression trees on bootstrapped samples from the training set and then taking an average over all predictions from all regression trees (Hastie et al., 2009). Each regression tree was generated by successively splitting into two the predictor space spanned by p predictors. At each binary split m randomly selected predictors were considered, and then the predictor and the cutpoint were selected further such that the sum of squared residuals was minimized within each of resulting partitions (James et al., 2013). In this research R library "randomForest" (Liaw and Wiener, 2002) was used for modelling data with random forest: 500 regression trees were generated and $m = p/3$ were selected at each binary split.

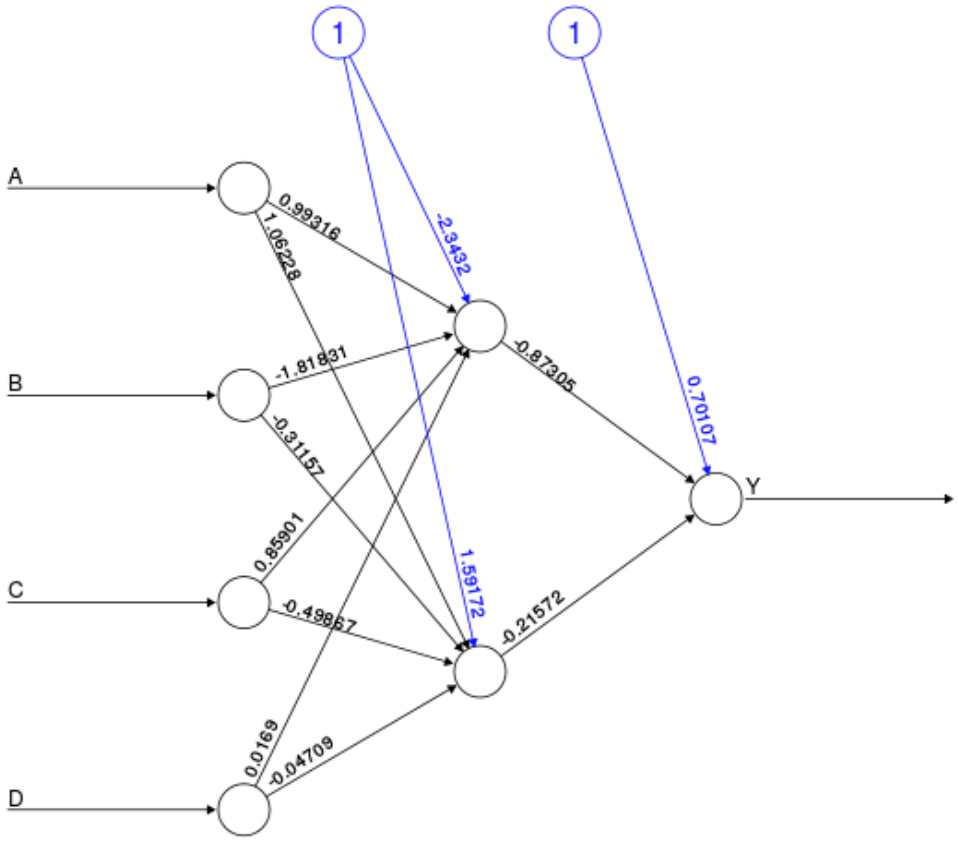

Figure 2.4: An example of a neural network with four input variables (A, B, C, D), an output neuron (Y) and two neurons in a hidden layer.

2.5.3 Artificial neural networks

Artificial neural networks (ANN) is a non-linear statistical model, whose structure consists of layers of neurons (Figure 2.4). ANN in this work is the multi-layer perceptron, in which each neuron in a subsequent layer is connected to all neurons from the previous layers via synapses (Zhang et al., 1998). Each synapse has a weight and the output y of each neuron is given by:

$$y = f(w_0 + \sum_{i=1}^{p} w_i x_i)$$

where w_0 is a bias term at each layer, w_i denotes a synaptic weight and x_i is a neuron output from the preceeding layer, and p is the number of neurons in the preceeding layer. Multi-layer ANN are well-suited to model non-linear data due to a non-linear activation function f applied to each neuron output (Bishop et al., 1995). In this research the hyperbolic tangent: $tanh\ x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ was used to transform the output x of each neuron to be between [-1,1].

The tuning parameters that define the model architecture of ANN are the number of layers and the number of neurons in each layer. Since a continuous function can be estimated with ANN with two hidden layers (Kurkova, 1992), the multi-layer perceptron in this work is ANN(L_1, L_2) where L_1 (L_2) denotes the number of neurons in the first (second) layer. A larger number of neurons in each layer increases the complexity of ANN and computational cost; therefore, the number of neurons in each layer was restricted to be no more than 10. Additionally, since ANN results are strongly influenced by starting weights (Hastie et al., 2009), 30 models with random starting weights drawn from the standard normal distribution were trained for ANN and both regression and forecast results were averaged across these 30 model outputs. Modeling response variables with artificial neural networks in this work was achieved by using the R library "neuralnet" (Günther and Fritsch, 2010).

2.5.4 Support vector regression

Support vector regression (SVR) belongs to a class of non-linear statistical models (Vapnik, 2000). SVR transforms the predictor space into a space with a higher number of dimensions using a non-linear kernel (Moguerza and Muñoz, 2006; Basak et al., 2007). ϵ-regression with the radial basis kernel given by $K(x_i, x_j) = e^{-\gamma||x_i - x_j||^2}$ with $\gamma > 0$ was employed in this research using R library "e1071" (Chang and Lin, 2011). The objective of ϵ-regression is to fit data with a function $f(x) =< w, x > +b$ by minimizing the cost function

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$subject \quad to \qquad \begin{cases} y_i - < w, x > -b \leq \epsilon + \xi_i \\ < w, x > +b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

where $< ., . >$ is the inner product and $||\cdot||^2$ is the Euclidean norm, C > 0 is a trade-off parameter between the flatness of the function f and the largest deviation greater than ϵ,

and ξ_i, ξ_i^* are introduced to deal with the optimization constraints (Smola and Schölkopf, 2004). The ϵ parameter was set at value of 0.1, while γ and C parameters were tuned as a part of model training with a grid search of (γ, C) where γ values are {0.1, 0.5, 1} and C values are {0.1,1,10}. In contrast to ANN, SVR does not have issues associated with local minima and therefore produces the unique solution.

2.6 Measures of predictive accuracy

The predictive performance measures used in this work are the root mean square error (RMSE) and the correlation coefficient (ρ).

- The test RMSE, which is a measure of deviation between forecasted values and observations in the test set, is determined as the square root of the sum of squared differences between the observations y_i and forecasted values $\hat{y}_i$ in the fixed test set such as the 2005-2011 time period

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{i}} - \hat{y}_i)^2} \qquad \forall i = 1, ..., n$$

  where n is the total number of values in the test set.

- The test ρ is the correlation between the observations y_i and forecasted $\hat{y}_i$ values over the test set.

$$\rho = \frac{\mu_{y\hat{y}} - \mu_y \mu_{\hat{y}}}{\sigma_y \sigma_{\hat{y}}}$$

  where μ denotes the average and σ is the standard deviation.

2.7 Optimal model architecture

The model architecture refers to the presence of the tuning parameters that define statistical model structure and complexity. To determine the optimal model architecture in order to generate reliable forecasts, k-fold cross-validation is used to tune model parameters. The cross-validation RMSE is generated for different values of a tuning parameter for a given

class of statistical models. The tuning parameter that results in the lowest cross-validation RMSE is considered optimal and the statistical model with such tuning parameter is used further to generate the forecast over the test set. In contrast to BM12, the optimal model architecture in this work is based on minimizing $RMSE_{test}$ instead of maximizing ρ_{test}. The correlation coefficient ρ was retained to compare the results against the findings by BM12.

MLR does not have tuning parameters. For RF, the tuning parameter m, a number of randomly selected predictors at each binary split, was set at the default value of $p/3$ (Liaw and Wiener, 2002). Similar to BM12, the model architecture for ANN(L_1, L_2) was established by 5-fold cross-validation using the R library "caret" (Kuhn, 2008). Likewise, the support vector regression parameters γ and C were also tuned with 5-fold cross-validation. The optimal model architecture ANN(L_1, L_2) and the optimal values of the SVR tuning parameters (γ, C) for response variables are shown in Table 3.1

2.8 Assessment of predictor importance

Since linear and non-linear statistical models model the relationship between predictors and response variables differently, the importance of each predictor may differ for various statistical models. Instead of performing the variable selection, all predictors were used for regression and the importance of each predictor was determined by conducting a permutation-based method (Ruß and Brenning, 2010). The permutation-based method implemented in R library "sperrorest" (Brenning, 2012) enables one to determine the importance of each predictor among statistical models by performing cross-validation. First, the cross-validation mean square error (MSE) is determined:

$$MSE = \frac{1}{kn}\sum_{i=1}^{n}\sum_{j=1}^{k}(y_{ij} - \hat{y_{ij}})^2 \qquad \forall i = 1, ..., n; j = 1, ..., k$$

where y_{ij} are observations and $\hat{y}_{ij}$ are forecasted values in each test fold, k is the total number of values in a given test fold and n is the number of folds. Second, r permutations of the values of the predictor are performed on each test set and the resulting MSE computed for each permutation. Then the average MSE is computed across all n folds and r permutations to obtain "permuted" cross-validation MSE. Finally, cross-validation MSE is subtracted from "permuted" cross-validation MSE to arrive at ΔMSE. The larger ΔMSE, the more important the predictor is for modeling the response variable with a statistical model. In this research leave-one-out cross-validation was performed over the

training set between July 1, 1980 and June 30, 2005 with each fold corresponding to an annual subset e.g. fold1 is the time period between July1, 1980 and June 30, 1981. For each predictor its values were randomly permuted over the test set 1000 times and the resulting ΔMSE was determined.

Chapter 3

Results

This section outlines the regression and forecast results obtained from the statistical models. The importance of each predictor for generating 10-day and 20-day forecasts of the northern polar stratospheric temperature and geopotential anomalies is quantified using a permutation-based method. The sensitivity of forecast skill to a hindcast test period is assessed. The impact of a strong SSW on the forecast accuracy is determined by replacing the values of response variables during a major SSW event with the climatological mean daily values and assessing the resulting change in predictive performance measures. The effect of extending the duration of the training time period on the predictive performance is also established.

3.1 Assessment of model predictive skill/performance

The statistical models were trained over the time period between July 1, 1980 and June 30, 2005 and forecasts were generated for the time period between July 1, 2005 and June 30, 2011. Similar to BM12, machine learning models fit data well over the training set as indicated by a much larger ρ_{train} and a smaller $RMSE_{train}$ in comparison to the MLR model (Table 3.1); however, the predictive performance measures over the test set are much worse indicating data overfitting. Among all models SVR overfits the most and consequently generates forecasts with poor accuracy.

The time series of observations and 10-day forecasts for the 2005-2011 time period are presented in Figure 3.1. For the 10-day temperature forecasts, both ANN and MLR produce the smallest $RMSE_{test}$ among all statistical models: 5.19 for ANN and 5.13

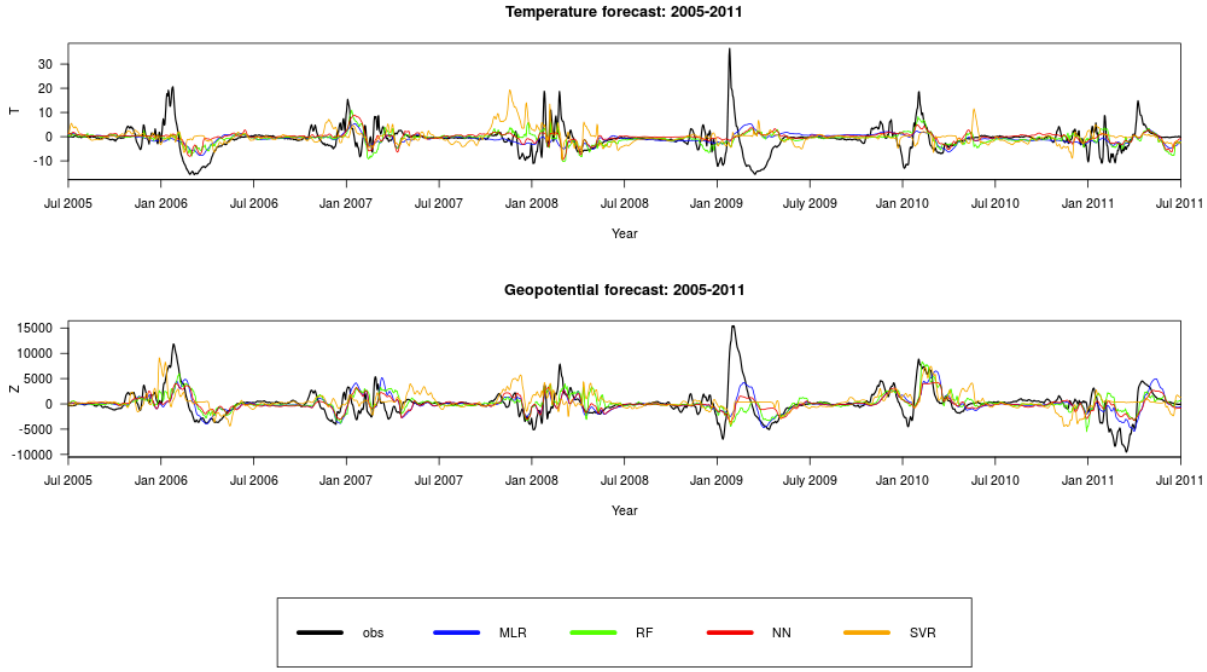

Figure 3.1: The observations (solid black line) and 10-day forecasts from statistical models for the time period between July 1, 2005 and June 30, 2011 for temperature T (units in K) and geopotential height Z (units in m) response variables.

for MLR (Table 3.1). By the Ockham's Razor principle (Ariew, 1976) when two models produce the same skill, a simpler model is preferred, and therefore MLR with ρ_{test} of 0.22 was preferable to ANN for modeling temperature anomalies. For 10-day geopotential height forecasts, based on the lowest $RMSE_{test}$ MLR outperformed machine learning models with ρ_{test} of 0.48. Among machine learning models ANN had the best predictive performance with ρ_{test} of 0.38; however, $RMSE_{test}$ for ANN was 4% larger than $RMSE_{test}$ for the MLR model. Such results were in contrast to BM12 results, which showed that ANN with ρ_{test} of 0.42 had outperformed MLR with ρ_{test} of 0.31 in forecasting geopotential height anomalies.

For the 20-day temperature forecasts, ρ_{test} was 0.24 and 0.23 for MLR and ANN, respectively. Therefore, the accuracy of the 10-day and 20-day temperature forecasts was similar. For the 20-day geopotential height forecasts, MLR showed the best predictive

Table 3.1: Training and test correlation coefficient (ρ) and root mean square error ($RMSE$) for (a) temperature (units in K) and (b) geopotential height (units in m).

a) Temperature

	10-day forecast				20-day forecast			
Model	MLR	RF	ANN	SVR	MLR	RF	ANN	SVR
Parameters			(10,4)	(1,10)			(10,10)	(1,10)
ρ_{train}	0.39	0.97	0.76	0.99	0.39	0.96	0.79	0.98
ρ_{test}	**0.22**	0.18	0.23	-0.05	**0.24**	0.23	0.22	0.01
$RMSE_{train}$	4.48	1.48	3.34	0.73	4.48	1.51	3.17	0.94
$RMSE_{test}$	**5.13**	5.33	5.19	6.14	**5.09**	5.18	5.22	5.51

b) Geopotential height

	10-day forecast				20-day forecast			
Model	MLR	RF	ANN	SVR	MLR	RF	ANN	SVR
Parameters			(8,5)	(1,10)			(10,8)	(1,10)
ρ_{train}	0.63	0.96	0.72	0.99	0.47	0.97	0.69	0.99
ρ_{test}	**0.48**	0.32	0.38	0.16	**0.34**	0.13	0.27	-0.04
$RMSE_{train}$	1983	742	1800	257	2255	7.35	1923	315
$RMSE_{test}$	**2460**	2709	2564	2990	**2636**	2923	2674	3033

performance among statistical models in this research. However, while the accuracy of the 10-day and 20-day temperature forecasts was very similar, the forecast accuracy for geopotential height deteriorated going from the 10-day to 20-day forecasts for all models (Table 3.1). $RMSE_{test}$ produced by MLR increased by 7% and ρ_{test} decreased from 0.48 to 0.34.

Although the 10-day and 20-day temperature ρ_{test} results were comparable with the correlation results obtained by BM12 using MLR and ANN, the forecast accuracy of the temperature forecasts was still very low. For geopotential height, the forecast accuracy diminished with the duration of forecasts. For the 10-day forecast MLR ρ_{test} of 0.48 was larger than ANN ρ_{test} of 0.42 as reported by BM12. Finally, same as in BM12, none of the statistical models in this research was able to accurately forecast very strong wintertime northern polar stratospheric anomalies such as 2009 SSW (Figure 3.1).

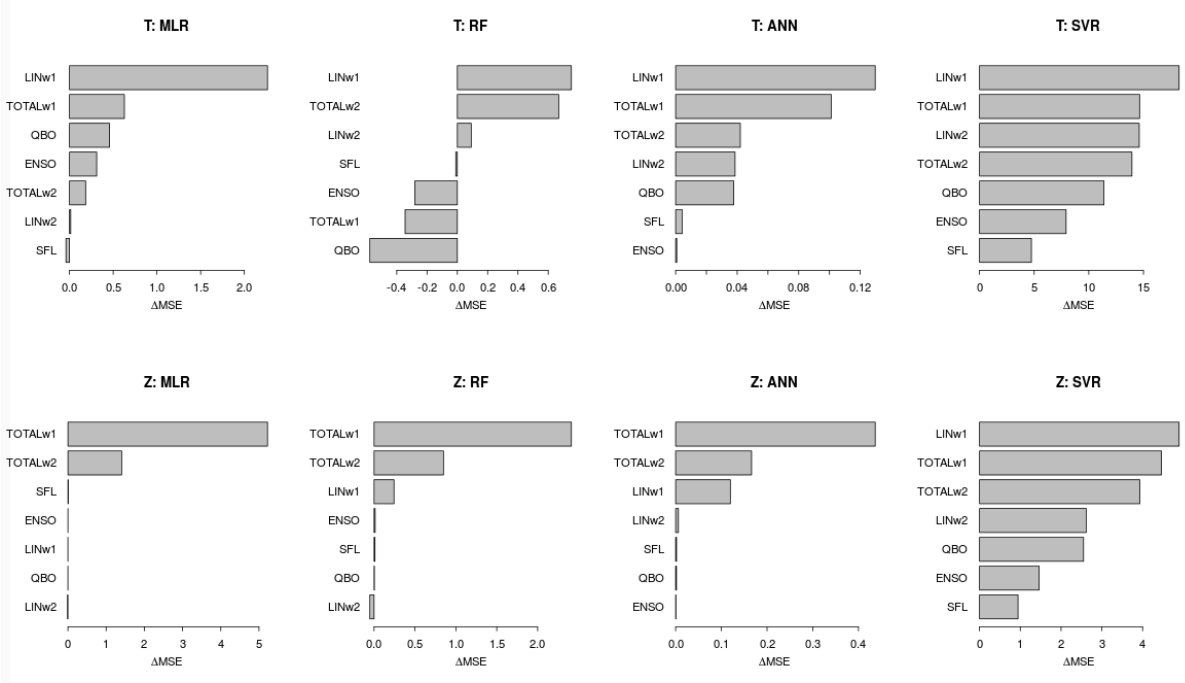

Figure 3.2: The importance of each predictor for generating the 10-day temperature T (units in K^2) and geopotential height Z (units in $10^6 m^2$) forecasts using statistical models.

3.2 Assessment of predictor importance

The importance of each predictor was determined by quantifying ΔMSE, a change in cross-validation MSE as a result of multiple permutations of the predictor's values over the test folds while performing cross-validation.

For 10-day temperature forecasts, LINw1 was shown by statistical models to be an important predictor for modeling temperature anomalies (Figure 3.2). The importance of the LINw1 predictor was most apparent for MLR as ΔMSE was the largest among all predictors. For RF, ΔMSE was negative for some predictors. The decrease in cross-validation MSE for such variables was likely due to random variable selection at each binary split during model construction for RF. Hence, only those predictors with positive ΔMSE values that were larger than the absolute value of the largest negative ΔMSE were considered important. Because ΔMSE for LINw1 was larger than the absolute value of ΔMSE for QBO, LINw1 was considered an important predictor for modeling temperature variability

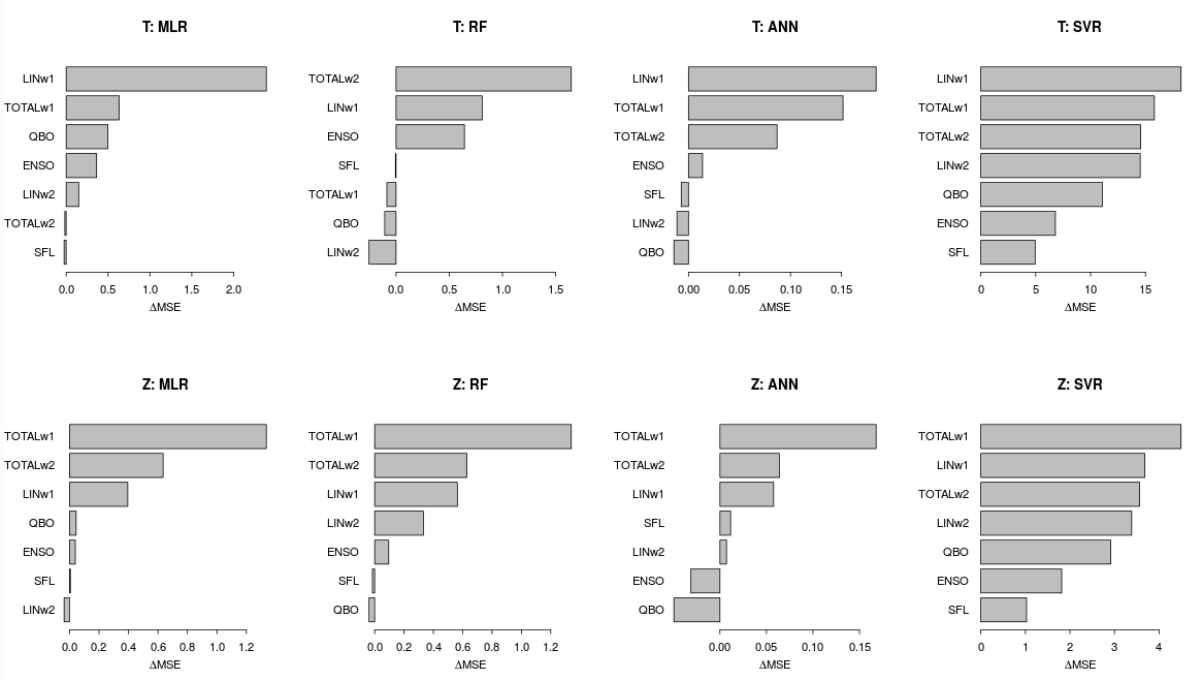

Figure 3.3: Same as in Figure 3.2, but for the 20-day forecasts.

with RF. For ANN, LINw1 was an important predictor followed by TOTALw1. For SVR, cross-validation MSE increased in response to permuting the values of each predictor and the largest ΔMSE was again observed for LINw1. In contrast to temperature, TOTALw1 and TOTALw2 are important for modeling geopotential height anomalies (Figure 3.2). Across all models, the QBO, ENSO, SFL predictors appear to be less important for generating 10-day geopotential height forecasts than predictors representing an upward heat flux entering the stratosphere.

The importance of each predictor for generating the 20-day temperature and geopotential height forecasts was summarized in Figure 3.3. With an exception of ENSO and TOTALw2, the optimal lags between temperature response variable and predictors exceeded the minimum lag of 20 days, and therefore the importance of each predictor for generating the 10-day and 20-day temperature forecasts were very similar. LINw1 was shown to be the leading predictor highlighting the importance of linear interference (Figure 3.3). For geopotential height, the increase in the duration of forecasts from 10 to 20 days led to a decreased importance of the heat fluxes for all models (Figure 3.3) and as

a result a much worse accuracy of forecasts. TOTALw1 was still the dominant predictor; however, the importance of TOTALw1 was diminished in comparison with the importance of the TOTALw1 for the 10-day forecasts. Such decreased TOTALw1 importance emphasized the sensitivity of the TOTALw1 predictor to extending the duration of forecasts and, as a result, the accuracy of the 20-day forecasts was much lower than the accuracy of the 10-day forecasts.

In summary, forecasting short-term northern polar stratospheric variability was improved when the predictors representing the upward heat flux entering the stratosphere were used: LINw1 for temperature and TOTALw1 for geopotential height. While there was no decrease in forecast skill going from the 10-day to 20-day temperature forecasts, for the 20-day forecast of geopotential height anomalies the decreased importance of TOTALw1 led to a large deterioration in predictive performance for all statistical models.

3.3 Sensitivity to hindcast time period

To determine whether the accuracy of forecasts varies for a different hindcast time period, the accuracy of 10-day forecasts over the period between July 1, 2011 and March 31, 2015 was assessed in addition to forecast skill over the time period between July 1, 2005 and June 30, 2011. The forecast skill over the 2011-2015 time period increased for all statistical models. For temperature, MLR had better predictive performance than the machine learning models and had ρ_{test} of 0.35. For geopotential height, the increase in ρ_{test} for MLR was modest: ρ_{test} was 0.48 for the 2005-2011 time period and 0.51 for the 2005-2011 time period. In contrast, both RF and ANN produced $RMSE_{test}$ results that were 4% and 2% lower than MLR with ρ_{test} of 0.57 and 0.55 respectively. Increased predictive performance from machine learning models over MLR indicated that for geopotential height the forecast skill was sensitive to a different hindcast period.

One of the possible explanations for an increased predictive performance of the machine learning models over the 2011-2015 time period (Figure 3.4) is the absence of very strong SSWs such as 2009 SSW in the 2005-2011 time period. Could the 2009 SSW have affected the forecast results and have led to worse forecast accuracy over the 2005-2011 test period? To determine the impact of the 2009 SSW event on the 10-day forecast results, an experiment was conducted with 2009 January and February daily values of response variables being replaced with the January and February climatological mean daily values computed over the time period between 1980 and 2005. As a result, both ρ_{test} and $RMSE_{test}$ for all models improved. The best predictive performance came from MLR with ρ_{test} of 0.27 and 0.52 for temperature and geopotential height respectively. The increase in forecast skill

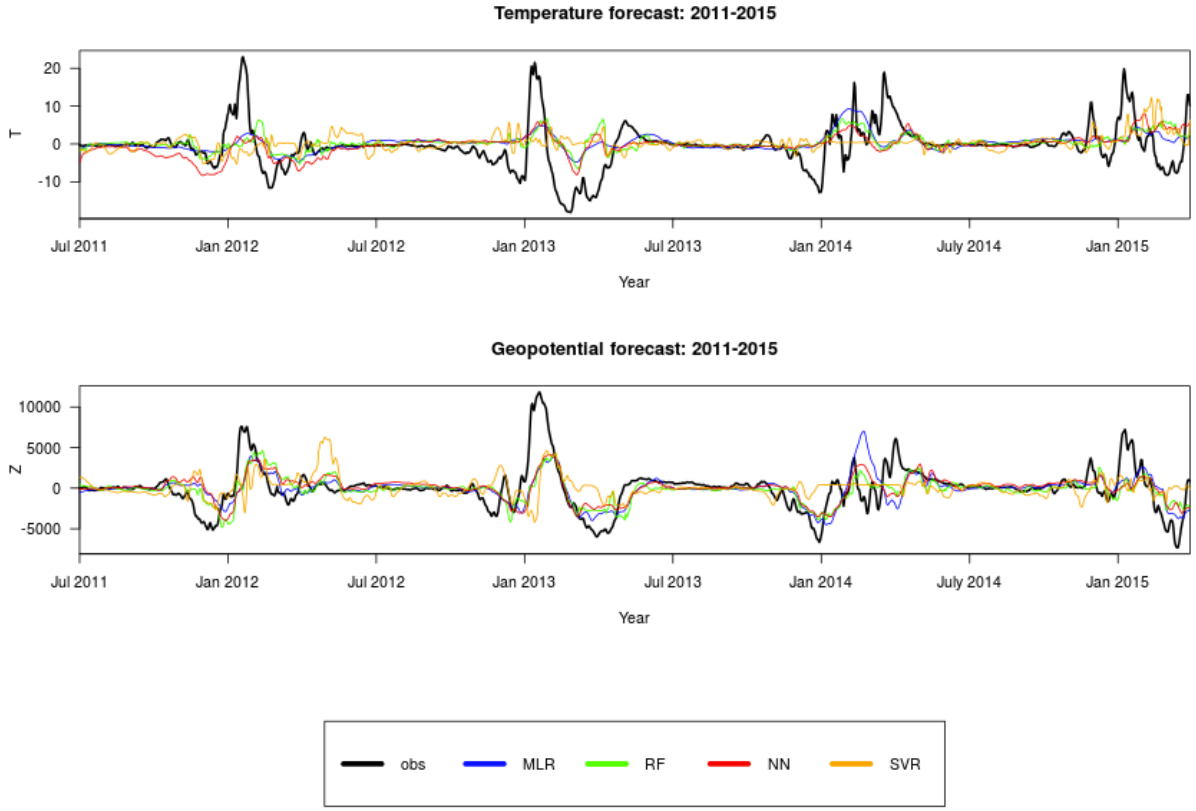

Figure 3.4: Same as in Figure 3.1, but the time period is between July 1, 2011 and March 31, 2015.

after replacing 2009 SSW values with the climatological mean daily values supported the claim that the presence of strong wintertime anomalies such as 2009 SSW led to a poorer forecast skill over the 2005-2011 time period as compared to the 2011-2015 time period.

Furthermore, an increased predictive performance over the 2011-2015 time period raised an additional question as to whether forecast results over the 2011-2015 time period would improve even further if the training set between 1980 and 2005 was extended by 6 years i.e. the 1980-2011 time period. To establish if a larger training set would improve the predictive performance of the 10-day forecasts of temperature and geopotential height anomalies, the statistical models were trained over the period between July 1, 1980 and June 30, 2011 and then tested over the time period between July 1, 2011 and March 31, 2015. Increasing the

time interval of the training set resulted in less overfitting; however, the predictive accuracy was very similar to the 10-day forecasts generated with the training period between July 1, 1980 and June 30, 2005. For temperature, MLR had $RMSE_{test}$, which was 3% lower than $RMSE_{test}$ from the machine learning models, and the largest ρ_{test} of 0.31. For geopotential height, however, ANN had the lowest $RMSE_{test}$ among statistical models and produced ρ_{test} of 0.57, which was similar to ρ_{test} of 0.57 from RF and 0.55 from ANN when a smaller training set had been used. Therefore, increasing the size of the training set did not lead to a better predictive performance of statistical models over the 2011-2015 time period.

Chapter 4

Discussion and Conclusions

This section contains the discussion of results and conclusions, explains limitations encountered in this research, as well as suggests the possible directions for future work.

Prior research by BM12 indicated that northern polar stratospheric temperature and geopotential height anomalies could be forecasted with statistical models. The goal of this research was to demonstrate that short-term forecasts of northern polar stratospheric temperature and geopotential height anomalies could be achieved with no loss in predictive accuracy while addressing two limitations of the BM12 approach: high computational cost and a lack of lagged predictors for generating forecasts. In addition to the climate indices representing the QBO, ENSO and solar variability, the upward heat flux of wavenumbers 1 and 2 entering the stratosphere were used as predictors. Multiple linear regression and three machine learning models, namely random forest, artificial neural networks, and support vector regression were employed to generate the 10-day and 20-day temperature and geopotential height forecasts. The predictive accuracy of the 10-day forecasts generated with this approach was similar for temperature and slightly higher for geopotential height than the forecast results reported by BM12. In addition to investigating the short-term predictability of northern polar stratospheric variability, this research sought to address four research objectives outlined in Section 1.4.

4.1 Relevance of heat flux predictors

The first objective of this research was to explore the importance of meridional wave heat fluxes as predictors in forecasting northern polar stratospheric variability as opposed to

the QBO, ENSO and the 11-year solar cycle. The northern polar stratospheric variability can be forecasted to a certain extent using linear and non-linear statistical models with the given set of lagged predictors. For temperature, the forecast skill over the time period between July 1, 2005 and June 30, 2011 remains poor as indicated by ρ_{test} of 0.22 while for geopotential height ρ_{test} of 0.48 for MLR is larger than the best ρ_{test} of 0.42 using ANN as reported by BM12. Thus, an improvement in the accuracy of short-term northern polar stratospheric geopotential height forecasts was achieved by adding meridional wave heat fluxes as the predictors. While LINw1, the linear component of the total meridional wave heat flux of wavenumber 1, was shown to be an important predictor for modeling short-term northern polar stratospheric temperature, the use of TOTAL heat fluxes, especially TOTALw1, as predictors greatly improves the accuracy of short-term geopotential height forecasts. Such results support the prior research (Polvani and Waugh, 2004; Nakagawa and Yamazaki, 2006; Charlton and Polvani, 2007; Fletcher and Minokhin, 2015) that the meridional wave heat flux with wavenumbers 1 and 2 entering the stratosphere have the potential as predictors to improve the accuracy of northern polar stratospheric temperature and geopotential height forecasts.

4.2 Predictive performance of statistical models

The second objective was to determine whether machine learning models exhibit higher predictive performance than multiple linear regression in forecasting northern polar stratospheric variability. In contrast to BM12, machine learning models in this research do not exhibit a better forecast accuracy than MLR for forecasting northern polar stratospheric geopotential height anomalies. This result can be explained by the lack of interaction between the predictors and a much larger importance of the predictors representing heat flux wave-1 propagating into the stratosphere. For both 2005-2011 and 2011-2015 test sets, MLR maintained a better predictive performance than the machine learning models for temperature anomalies. However, the choice of the test set influenced the model selection for forecasting geopotential height anomalies. RF has better predictive performance than MLR over the 2011-2015 test set and therefore may seem to be more suitable for generating 10-day geopotential height forecasts in the future. However, the fact than RF could not outperform MLR consistently over the two test sets indicates that MLR should be preferred since its ρ_{test} over both tests were very similar: 0.48 and 0.51.

It is possible that in this research the model architecture of machine learning models may not have been optimal. The range of the tuning parameters was restricted to reduce the computational cost. For instance, for ANN the number of neurons in the first and second

layers was set to be less than or equal to 10. Moreover, the repeated cross-validation with a number of folds larger than five might point to different values of tuning parameters for machine learning models. Although increasing the complexity of the model architecture of machine learning models might result in a better predictive performance in comparison to MLR, such increased complexity will certainly result in a higher computational cost, which would become a limitation of the approach used in this research.

4.3 Sensitivity of forecast skill to time lags

The third objective was to investigate the sensitivity of the accuracy of forecasts with respect to time lags and to test whether there is a deterioration in the accuracy of the 20-day forecasts as opposed to the 10-day forecasts. A limitation of using meridional wave heat fluxes as predictors is that the predictability of heat fluxes is relatively short. While for temperature the forecast skill did not decrease for the 10-day and 20-day forecasts, the forecast skill is considered rather poor. For geopotential height, the forecast accuracy is sensitive to the duration of forecasts. The forecast skill of the 20-day geopotential height forecasts decreased due to a lower contribution of the TOTALw1 heat flux to overall forecast skill. A short duration of northern polar stratospheric forecasts is certainly disadvantageous and further attempts to improve forecast accuracy over a longer time period should be undertaken.

The method of computing optimal lags between the response variables and predictors may have affected the quality of regression and forecast results. Despite the ease of computing linear lags based on the largest absolute correlation between the response variables and lagged predictors, such method for determining the optimal lags may not lead to a high model predictive performance and therefore the possible use of non-linear lags based on minimizing cross-validation RMSE should be explored. However, in agreement with BM12, the computational cost associated with pursuing such approach to determine optimal non-linear lags will be higher than the approach based on linear lags.

4.4 Predictability of SSW

The fourth objective was to investigate if SSW, which is an example of extreme polar stratospheric events, can be forecasted with statistical models using the proposed set of predictors. Despite the improvement in accuracy of forecasts as a result of using the heat fluxes as predictors, forecasting strong SSWs remains a challenge. One of the explanations

as to why SSWs are hard to predict is the necessary "preconditioning" of the SSWs when the resonant wave enhancement and the internal stratospheric variability play a crucial role in the evolution of SSWs (Stan and Straus, 2009; Albers and Birner, 2014). Incorporating the resonant enhancement of planetary waves propagating from the troposphere may lead to a better prediction of SSWs. Furthermore, none of the statistical models with the given set of predictors in this research was able to accurately forecast the very large wintertime fluctuations in the temperature and geopotential time series. Stan and Straus (2009) argued that northern polar stratospheric variability was non-stationary during winters. If predictor characteristics such as the mean and variance vary greatly over time, then such variability may not be adequately captured with the statistical models used in this research. This research did not employ a non-stationary statistical model for modeling and forecasting northern polar stratospheric variability. Consequently, it remains to be investigated if the use of non-stationary models such as state-space models can lead to more accurate forecasts and an improvement in the prediction of SSWs. State-space models have shown promising results when used to model stratospheric ozone trends (Laine et al., 2014),

4.5 Future work

Future work can address the limitations of this research and investigate whether the forecast accuracy can be improved further and whether the duration of forecasts can be extended to meet the demands of decision makers. First, modeling resonant enhancement of planetary waves as an predictor may lead to improved forecast results and a better predictability of SSWs. Second, it remains to be examined if the accuracy of forecasts will improve when non-stationary statistical models such as state-space models are used. Third, another reanalysis dataset for modeling and forecasting northern polar stratospheric variability can be considered. The ERA-Interim dataset is not updated regularly and therefore continuous 10-day forecasts are not possible. Instead, other reanalysis datasets such as the Japanese 55-year Reanalysis[1](JRA-55) could be used to generate continuous forecasts and validate the regression results obtained from the ERA-Interim dataset. Finally, the stratospheric forecasts have the potential to be used to tune parameters in the temperature pricing model (Alaton et al., 2002). Hence, the incorporation of northern polar stratospheric forecasts into weather derivative pricing models remains to be explored in future work.

[1]https://reanalyses.org/atmosphere/overview-current-reanalyses/#JRA-55

References

Alaton, P., Djehiche, B., and Stillberger, D. (2002). On modelling and pricing weather derivatives. *Applied mathematical finance*, 9(1):1–20.

Albers, J. R. and Birner, T. (2014). Vortex preconditioning due to planetary and gravity waves prior to sudden stratospheric warmings. *Journal of the Atmospheric Sciences*, 71(11):4028–4054.

Andrews, D. G., Holton, J. R., and Leovy, C. B. (1987). *Middle atmosphere dynamics*. Number 40. Academic press.

Ariew, R. (1976). Ockham's razor: A historical and philosophical analysis of ockham's principle of parsimony.

Baldwin, M. P. and Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, 294(5542):581–584.

Basak, D., Pal, S., and Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10):203–224.

Bishop, C. M. et al. (1995). Neural networks for pattern recognition.

Blume, C. and Matthes, K. (2012). Understanding and forecasting polar stratospheric variability with statistical models. *Atmospheric Chemistry and Physics*, 12(13):5691–5701.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The r package sperrorest. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 5372–5375. IEEE.

Butler, A. H. and Polvani, L. M. (2011). El niño, la niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophysical Research Letters*, 38(13).

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Charlton, A. J. and Polvani, L. M. (2007). A new look at stratospheric sudden warmings. part i: Climatology and modeling benchmarks. *Journal of Climate*, 20(3):449–469.

Charlton, A. J., Polvani, L. M., Perlwitz, J., Sassi, F., Manzini, E., Shibata, K., Pawson, S., Nielsen, J. E., and Rind, D. (2007). A new look at stratospheric sudden warmings. part ii: Evaluation of numerical model simulations. *Journal of Climate*, 20(3):470–488.

Charron, M., Polavarapu, S., Buehner, M., Vaillancourt, P., Charette, C., Roch, M., Morneau, J., Garand, L., Aparicio, J. M., MacPherson, S., et al. (2012). The stratospheric extension of the canadian global deterministic medium-range weather forecasting system and its impact on tropospheric forecasts. *Monthly Weather Review*, 140(6):1924–1944.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Deo, R. C. and Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern australia. *Atmospheric Research*, 153:512–525.

Fletcher, C. G. and Minokhin, I. (2015). Linear interference and the northern annular mode response to el niño and climate change. *Climate Dynamics*, pages 1–15. ISSN: 0930–7575, 1432–0894.

Garfinkel, C., Butler, A., Waugh, D., Hurwitz, M., and Polvani, L. M. (2012). Why might stratospheric sudden warmings occur with similar frequency in el niño and la niña winters? *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D19).

Garfinkel, C. and Hartmann, D. (2008). Different enso teleconnections and their effects on the stratospheric polar vortex. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113(D18).

Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., Bovensmann, H., Burrows, J., Degenstein, D., Froidevaux, L., and Thompson, A. (2014). Stratospheric ozone trends and variability as seen by sciamachy from 2002 to 2012. *Atmospheric Chemistry and Physics*, 14(2):831–846.

Gerber, E., Orbe, C., and Polvani, L. M. (2009). Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophysical Research Letters*, 36(24).

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., et al. (2010). Solar influences on climate. *Reviews of Geophysics*, 48(4).

Günther, F. and Fritsch, S. (2010). neuralnet: Training of neural networks. *The R Journal*, 2(1):30–38.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2. Springer.

Hsieh, W. W. (2009). *Machine learning methods in the environmental sciences: neural networks and kernels*. Cambridge university press.

Hsieh, W. W., Wu, A., and Shabbar, A. (2006). Nonlinear atmospheric teleconnections. *Geophysical research letters*, 33(7).

Hull, J., Treepongkaruna, S., Colwell, D., Heaney, R., and Pitt, D. (2013). *Fundamentals of futures and options markets*. Pearson Higher Education AU.

Ineson, S. and Scaife, A. (2009). The role of the stratosphere in the european climate response to el niño. *Nature Geoscience*, 2(1):32–36.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Kerkhoff, C., Künsch, H. R., and Schär, C. (2014). Assessment of bias assumptions for climate models. *Journal of Climate*, 27(17):6799–6818.

Kuhn, M. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5):1–26.

Kurkova, V. (1992). Kolmogorov's theorem and multilayer neural networks. *Neural networks*, 5(3):501–506.

Laine, M., Latva-Pukkila, N., and Kyrölä, E. (2014). Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmospheric Chemistry and Physics*, 14(18):9707–9725.

Liaw, A. and Wiener, M. (2002). randomforest: Classification and regression with random forest. *R package*.

Lu, H., Baldwin, M. P., Gray, L. J., and Jarvis, M. J. (2008). Decadal-scale changes in the effect of the qbo on the northern stratospheric polar vortex. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113(D10).

Moguerza, J. M. and Muñoz, A. (2006). Support vector machines with applications. *Statistical Science*, pages 322–336.

Mukougawa, H. and Hirooka, T. (2004). Predictability of stratospheric sudden warming: A case study for 1998/99 winter. *Monthly weather review*, 132(7):1764–1776.

Nakagawa, K. I. and Yamazaki, K. (2006). What kind of stratospheric sudden warming propagates to the troposphere? *Geophysical research letters*, 33(4).

Naoe, H. and Shibata, K. (2010). Equatorial quasi-biennial oscillation influence on northern winter extratropical circulation. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D19).

Newman, P. A., Nash, E. R., and Rosenfield, J. E. (2001). What controls the temperature of the arctic stratosphere during the spring? *Journal of Geophysical Research: Atmospheres (1984–2012)*, 106(D17):19999–20010.

Ouzeau, G., Cattiaux, J., Douville, H., Ribes, A., and Saint-Martin, D. (2011). European cold winter 2009–2010: How unusual in the instrumental record and how reproducible in the arpege-climat model? *Geophysical Research Letters*, 38(11).

Polvani, L. M. and Waugh, D. W. (2004). Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *Journal of climate*, 17(18):3548–3554.

Randel, W. J., Garcia, R. R., Calvo, N., and Marsh, D. (2009). Enso influence on zonal mean temperature and ozone in the tropical lower stratosphere. *Geophysical Research Letters*, 36(15).

RCoreTeam (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Richter, J. H., Matthes, K., Calvo, N., and Gray, L. J. (2011). Influence of the quasi-biennial oscillation and el niño–southern oscillation on the frequency of sudden stratospheric warmings. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 116(D20).

Ruß, G. and Brenning, A. (2010). Spatial variable importance assessment for yield prediction in precision agriculture. In *Advances in Intelligent Data Analysis IX*, pages 184–195. Springer.

Sheshadri, A., Plumb, R. A., and Gerber, E. P. (2015). Seasonal variability of the polar stratospheric vortex in an idealized agcm with varying tropospheric wave forcing. *Journal of the Atmospheric Sciences*, (2015).

Sigmond, M., Scinocca, J., Kharin, V., and Shepherd, T. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, 6(2):98–102.

Smith, D. M., Scaife, A. A., and Kirtman, B. P. (2012). What is the current state of scientific knowledge with regard to seasonal and decadal forecasting. *Environ. Res. Lett*, 7(015,602).

Smith, K. L. and Kushner, P. J. (2012). Linear interference and the initiation of extratropical stratosphere-troposphere interactions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D13).

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

Stan, C. and Straus, D. M. (2009). Stratospheric predictability and sudden stratospheric warming events. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114(D12).

Thompson, D. W., Baldwin, M. P., and Wallace, J. M. (2002). Stratospheric connection to northern hemisphere wintertime weather: Implications for prediction. *Journal of Climate*, 15(12):1421–1428.

Trenberth, K. E. (1997). The definition of el nino. *Bulletin of the American Meteorological Society*, 78(12):2771–2777.

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S. D., Gerber, E., Harrison, R. G., Jackson, D. R., Kim, B.-M., Kuroda, Y., et al. (2014). The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quarterly Journal of the Royal Meteorological Society*.

Vapnik, V. (2000). *The nature of statistical learning theory*. Springer Science & Business Media.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R. (2014). A global perspective on cmip5 climate model biases. *Nature Climate Change*, 4(3):201–205.

Watson, P. A. and Gray, L. J. (2014). How does the quasi-biennial oscillation affect the stratospheric polar vortex? *Journal of the Atmospheric Sciences*, 71(1):391–409.

Zeng, Z., Hsieh, W., Shabbar, A., and Burrows, W. (2011). Seasonal prediction of winter extreme precipitation over canada by support vector regression. *Hydrology and Earth System Sciences*, 15(1):65–74.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.